REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS
BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
	AD A089835	

4. TITLE (and Subtitle)	5. TYPE OF REPORT & PERIOD COVERED
Projection Methods for Variational Inequalities with Application to the Traffic Assignment Problem	Technical
	6. PERFORMING ORG. REPORT NUMBER
	LIDS-P-1043

7. AUTHOR(s)	8. CONTRACT OR GRANT NUMBER(s)
Dimitri P. Bertsekas Eli M. Gafni	ARPA Order No. 3045/5-7-75 ONR/N00014-75-C-1183

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Massachusetts Institute of Technology Laboratory for Information and Decision Systems Cambridge, MA 02139	Program Code No. 5T10 ONR Identifying No. 049-383

11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE
Defense Advanced Research Projects Agency 1400 Wilson Boulevard Arlington, Virginia 22209	September 1980
	13. NUMBER OF PAGES
	36

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report)
Office of Naval Research Information Systems Program Code 437 Arlington, Virginia 22217	Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Symposium on Systems Analysis & Optimization, Univ. of Kentucky, June 1980.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

It is well known [1], [2], [3] that if $\overline{T}:R^n \rightarrow R^n$ is a Lipschitz continuous, strongly monotone operator and X is a closed convex set, then a solution $x^* \in X$ of the variational inequality

$$(x-x^*)'\overline{T}(x^*) \geq 0, \qquad \forall \ x \in X$$

can be found iteratively by means of the projection method

$$x_{k+1} = P_X[x_k - \alpha\overline{T}(x_k)], \qquad x_o \in X$$

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE
1 JAN 73

80 10 1 025

provided the stepsize α is sufficiently small. We show that the same is true if $\bar{T}$ is of the form

$$\bar{T} = A'TA,$$

where $A:R^n \rightarrow R^m$ is a linear mapping, provided $T:R^m \rightarrow R^m$ is Lipschitz continuous and strongly monotone, and the set X is polyhedral. This fact is used to construct an effective algorithm for finding a network flow which satisfies given demand constraints and is positive only on paths of minimum marginal delay or travel time.

Accession For	
NTIS GRA&I	✓
DTIC TAB	☐
Unannounced	☐
Justification	

By		
Distribution/		
Availability Codes		
Dist	Avail and/or Special	
A		

September 1980   39   LIDS-P-1043

PROJECTION METHODS FOR VARIATIONAL INEQUALITIES WITH APPLICATION TO THE TRAFFIC ASSIGNMENT PROBLEM[†]

Technical rept.,

by

Dimitri P. Bertsekas and Eli M. Gafni

Laboratory for Information and Decision Systems

Massachusetts Institute of Technology

Cambridge, Mass. 02139

410950

ABSTRACT

It is well known [1],[2],[3] that if $\bar{T}:R^n \to R^n$ is a Lipschitz
continuous, strongly monotone operator and X is a closed convex set,
then a solution $x^*\epsilon X$ of the variational inequality

$$(x-x^*)'\bar{T}(x^*) \geq 0, \qquad \forall\ x\epsilon X$$

can be found iteratively by means of the projection method

$$x_{k+1} = P_X[x_k - \alpha\bar{T}(x_k)], \qquad x_0\epsilon X,$$

provided the stepsize α is sufficiently small. We show that the same is
true if $\bar{T}$ is of the form

$$\bar{T} = A'TA,$$

where $A:R^n \to R^m$ is a linear mapping, provided $T:R^m \to R^m$ is Lipschitz
continuous and strongly monotone, and the set X is polyhedral. This
fact is used to construct an effective algorithm for finding a network
flow which satisfies given demand constraints and is positive only on
paths of minimum marginal delay or travel time.

1.   Introduction

We consider the problem of finding $x^{*} \epsilon X$ satisfying the variational inequality

$$(1) \qquad (x-x^{*})'A'T(Ax^{*}) \geq 0 \quad , \qquad\qquad \forall\ x\epsilon X$$

where X is a nonempty polyhedral subset of R^{n}, A is a given mxn matrix, and $T:R^{m} \rightarrow R^{m}$ is a nonlinear operator which is Lipschitz continuous and strongly monotone in the sense that there exist positive scalars L and λ such that for all y_1, y_2 in the set $Y = \{y|y=Ax, x\epsilon X\}$ we have

$$(2) \qquad |T(y_1) - T(y_2)| \leq L|y_1 - y_2| \quad ,$$

$$(3) \qquad [T(y_1) - T(y_2)]'\ (y_1 - y_2) \geq \lambda|y_1 - y_2|^2 \ .$$

In the relations above and throughout the paper all vectors are considered to be column vectors, and a prime denotes transposition. The standard norm in R^{n} is denoted $|\cdot|$, i.e. $|x| = (x'x)^{1/2}$ for all $x\epsilon R^{n}$.

We are interested in the projection algorithm

$$(4) \qquad x_{k+1} = P_X^S[x_k - \alpha S^{-1} A'T(Ax_k)], \qquad\qquad x_o \epsilon X$$

where $\alpha > 0$ is a stepsize parameter, S is a fixed positive definite symmetric matrix, and, for any $z\epsilon R^{n}$, $P_X^S(z)$ denotes the unique projection of z on the set X with respect to the norm $||\cdot||$ corresponding to S

$$(5) \qquad ||w|| = (w'Sw)^{1/2} \quad , \qquad\qquad \forall\ w\epsilon R^{n}.$$

The variational inequality (1) arises from the variational inequality

$$(6) \qquad (y - y^{*})'\ T(y^{*}) \geq 0 \quad , \qquad\qquad \forall\ y\epsilon Y$$

through the transformation

$$(7) \qquad y = Ax, \qquad\qquad Y = AX = \{y|y = Ax, x\epsilon X\} \quad .$$

It is possible to employ the projection algorithm

$$(8) \qquad y_{k+1} = P^Q_Y[y_k - \alpha Q^{-1}T(y_k)] , \qquad\qquad y_o \epsilon Y$$

for solving (6) where α is a positive stepsize parameter and Q is positive definite symmetric. It has been shown by Sibony [1] that if T is Lipschitz continuous and strongly monotone, and α is chosen sufficiently small then the sequence $\{y_k\}$ generated by (8) converges to the unique solution y^* of (6). The rate of convergence is typically linear although a superlinear convergence rate is possible in exceptional cases [13]. Strong monotonicity of the mapping T is an essential assumption for these results to hold. Our motivation for considering iteration (4) stems from the fact that in some cases projection on the set Y is very difficult computationally while projection on the set X through the transformation y = Ax [c.f. (7)] may be very easy. Under these circumstances if all other factors are equal, the projection method (4) is much more efficient than the method (8). This situation occurs for example in the application discussed in Section 3.

A potential difficulty with the transformation idea described above is that the mapping A'TA is not strongly monotone unless the matrix A'A is invertible. Thus convergence of iteration (4) is not guaranteed by the existing theory [1] - [3]. One of the contributions of this paper is to show that the convergence and rate of convergence properties of iteration (4) are satisfactory and comparable with those of iteration (8). These results hinge on the assumption that X is polyhedral, and it is unclear whether and in what form they hold if X is a general convex set.

In Section 3 we consider a classical traffic equilibrium problem arising in several contexts including communication and transportation networks, which can be modelled in terms of a variational inequality of

the form (1). A projection algorithm for solving this problem which is
essentially of the form (8) has been given by Dafermos [4]. Her algorithm
however operates in the space of link flows, and involves a projection iteration
which is very costly for large networks. We consider an alternative algorithm
which is basically of the form (4) and operates in the space of path flows.
Because the projection iteration can be carried out easily in this space
our algorithm is much more efficient. An algorithm which has several similarities
with ours has been proposed by Aashtiani [5] and has performed well in compu-
tational experiments. However, Aashtiani's algorithm cannot be shown to
converge in general. By contrast the results of Section 2 guarantee convergence
and linear rate of convergence for our method. There are also other methods
[7],[8],[10],[11],[12] for solving the special case of the traffic assignment
problem where T is a gradient mapping and there is an underlying convex
programming problem. Some of these methods [7], [8], [10] are of the projection
type. The algorithm of the present paper, however, seems to be the first
that can solve the general problem, is suitable for large networks, and
is demonstrably convergent.

In Section 4 we consider the generalized version of iteration (4),

$$x_{k+1} = P_X^{S_k} [x_k - \alpha \, S_k^{-1} A'T(Ay_k)], \qquad x_o \varepsilon X$$

where $\{S_k\}$ is a sequence of positive definite symmetric matrices with
eigenvalues bounded above and bounded away for zero. Projection algorithms
of this type include Newton's method for constrained minimization [14],[15], and
several network flow algorithms [5],[7] - [9]. Except for the case where
T is the gradient of a convex function, there are no convergence results
in the literature for this algorithm, even when A is the identity matrix.
We show that if care is taken in the way the matrices S_k are allowed to

change, then the resulting algorithm is convergent at a rate which is at least linear. We also provide a computational example involving a traffic assignment problem.

2. Projection Methods for Variational Inequalities

Let X^* be the set of all solutions of the variational inequality (1). We have that X^* is polyhedral and is given by

(9) $$X^* = \{x \in X \mid Ax = y^*\}$$

where y^* is the unique solution of the variational inequality (6). For any $x \in R^n$ we denote by $p(x)$ its unique projection on X^* with respect to the norm (5), i.e.

(10) $$p(x) = \arg \min \{\|x-z\| \mid z \in X^*\} .$$

We recall that projection on a convex set is a nonexpansive mapping (see e.g. [13]), so we have

(11) $$\|p(x_1) - p(x_2)\| \leq \|x_1 - x_2\| \quad , \quad \forall \, x_1, x_2 \in R^n .$$

We denote the mapping $A'TA$ by $\bar{T}$, i.e.

(12) $$\bar{T}(x) = A'T(Ax) \quad , \quad \forall \, x \in R^n .$$

With this notation we have

(13) $$(x - x^*)' \, \bar{T}(x^*) \geq 0, \quad \forall \, x \in X, \quad x^* \in X^* .$$

We also denote for all $x \in X$, $\alpha > 0$

(14) $$\hat{x}(x,\alpha) = P_X^S [x - \alpha S^{-1} \bar{T}(x)] .$$

With this notation the projection iteration (4) is written as

(15) $$x_{k+1} = \hat{x}(x_k, \alpha), \quad x_0 \in X.$$

It can be seen that we can also obtain x_{k+1} as the unique solution to the problem

(16) $$\text{minimize} \quad (x - x_k)' \, \bar{T}(x_k) + \frac{1}{2\alpha}(x - x_k)' \, S(x - x_k)$$
$$\text{subject to } x \in X \quad .$$

-6-

In view of (13) it is easy to see that for all solutions $x^* \epsilon X^*$ and $\alpha > 0$ we have

$$(17) \qquad \hat{x}(x^*,\alpha) = x^* = p(x^*) \ .$$

Thus if $x_k \epsilon X^*$ for some k the algorithm (15) essentially terminates.

Our main result is given in the following proposition.

<u>Proposition 1</u>: Assume that T is Lipschitz continuous and strongly monotone, and X is polyhedral.

a) There exist positive scalars $q(S)$ and $r(S)$ depending continuously on S, such that for all $x \epsilon X$ and $\alpha > 0$

$$(18) \qquad ||\hat{x}(x,\alpha) - p[\hat{x}(x,\alpha)]||^2 \leq [1 - 2\alpha q(S) + \alpha^2 r(S)] \ ||x - p(x)||^2.$$

b) There exists $\bar{\alpha} > 0$ such that for all $\alpha \epsilon (0,\bar{\alpha}]$ the sequence $\{x_k\}$ generated by iteration (15) converges to a solution x^* of the variational inequality (1). The rate of convergence is at least linear in the sense that for each $\alpha \epsilon (0,\bar{\alpha}]$ and $x_o \epsilon X$ there exist scalars β (depending on α) and q (depending on α and x_o) such that $q > 0$, $\beta \epsilon (0,1)$, and

$$||x_k - x^*|| \leq q \beta^k , \qquad\qquad k = 0,1,\dots$$

The proof of Proposition 1 relies on the following lemma, the proof of which is relegated to the appendix. The lemma is easy to conjecture in terms of geometrical arguments (see Figure 1).

<u>Lemma 1</u>: Assume the conditions of Proposition 1 hold. Let

$$(19) \qquad \tilde{X}^* = \{x \epsilon R^n | Ax = y^*\}$$

$$(20) \qquad \tilde{p}(x) = \arg \min \{||x - z|| \ |z \epsilon \tilde{X}^*\} \ .$$

Then $p(x) = p[\tilde{p}(x)]$ and there exists a positive scalar $\eta(S)$ depending continuously on S such that for all $x \epsilon X$

$$(21) \qquad ||x - \tilde{p}(x)||^2 \geq \eta(S)|| p(x) - \tilde{p}(x)||^2.$$

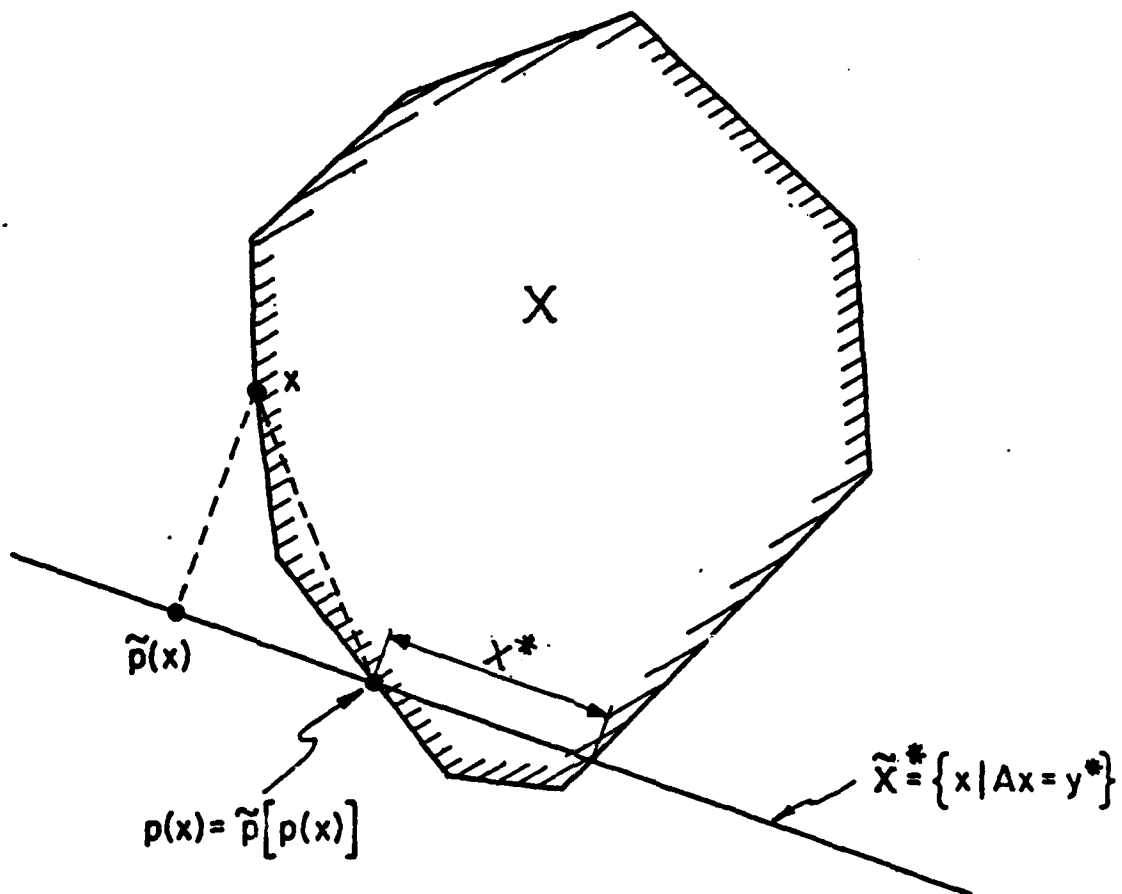**Figure 1:** Geometric Interpretation of Lemma 1

<u>Proof of Proposition 1</u>:   a) From (10) we obtain

(22) $||\hat{x}(x,\alpha) - p[\hat{x}(x,\alpha)]||^2 \leq ||\hat{x}(x,\alpha) - p(x)||^2$

Also, using (17) and (11) we have

(23) $||\hat{x}(x,\alpha) - p(x)||^2 = ||\hat{x}(x,\alpha) - \hat{x}[p(x),\alpha]||^2$

$\leq ||[x - p(x)] - \alpha S^{-1}[\bar{T}(x) - \bar{T}[p(x)]]||^2 = ||x - p(x)||^2$

$- 2\alpha[\bar{T}(x) - \bar{T}[p(x)]]'\,[x - p(x)] + \alpha^2||S^{-1}[\bar{T}(x) - \bar{T}[p(x)]]||^2 .$

In view of (22) and (23) it will be sufficient to show that there exists

positive scalars $q(S)$ and $r(S)$ depending continuously on S such that

(24a) $[\bar{T}(x) - \bar{T}[p(x)]]'\,[x - p(x)] \geq q(S)||x - p(x)||^2 \quad \forall\ x\epsilon X$

(24b) $||S^{-1}[\bar{T}(x) - \bar{T}[p(x)]]||^2 \leq r(S)||x - p(x)||^2 \qquad \forall\ x\epsilon X .$

To prove (24a) we first use (3) and (12) to obtain for all $x\epsilon X$ and $y = Ax$

(25) $[\bar{T}(x) - \bar{T}[p(x)]]'[x - p(x)] = [T(Ax) - T[Ap(x)]]'[Ax - Ap(x)]$

$= [T(y) - T(y^*)]'(y - y^*) \geq \lambda\ |y - y^*|^2 .$

We now consider next the range $R(A)$ and the nullspace $N(A)$ of A, and the

orthogonal complement

$$N(A)^+ = \{z\epsilon R^n|\ z'x\ = 0, \quad \forall\ x\epsilon N(A)\} .$$

When viewed as a mapping from $N(A)^+$ to $R(A)$, A is one-to-one and onto, and

since we are dealing with finite dimensional spaces we have that there

exists a positive scalar $\beta(S)$ depending continuously on S such that

(26) $|w|^2 \geq \beta(S)||z||^2, \quad \forall\ w\epsilon R(A),\ z\epsilon N(A)^+, \quad w = Az.$

Since for all $x\epsilon X$, $y = Ax$, and $\tilde{p}(x)$ defined by (20), we have

$$y - y^* = A[x - \tilde{p}(x)]$$

and

$$x - \tilde{p}(x) \epsilon N(A)^+$$

we obtain from (26)

(27) $$|y - y^*|^2 \geq \beta(S)||x - \tilde{p}(x)||^2.$$

Using the Pythagorean theorem and (21) we have

(28) $$||x - p(x)||^2 = ||x - \tilde{p}(x)||^2 + ||p(x) - \tilde{p}(x)||^2$$

$$\leq [1 + \frac{1}{\eta(S)}] \, ||x - \tilde{p}(x)||^2.$$

By combining (25), (27), and (28) we obtain

$$[\bar{T}(x) - \bar{T}[p(x)]]'[x - p(x)] \geq \lambda\beta(S)[1 + \frac{1}{\eta(S)}]^{-1}||x - p(x)||^2$$

so (24a) holds with

$$q(S) = \lambda\beta(S)[1 + \frac{1}{\eta(S)}]^{-1} \, .$$

To show (24b) we first use (2) and (12) to write

(29) $$||S^{-1}[\bar{T}(x) - \bar{T}[p(x)]]||^2 = [T(y) - T(y^*)]' \, AS^{-1}A'[T(y) - T(y^*)]$$

$$\leq \Lambda(S)|T(y) - T(y^*)|^2 \leq \Lambda(S) \, L^2|y - y^*|^2$$

where $\Lambda(S)$ is the largest eigenvalue of $AS^{-1}A'$.

By the same argument used to derive (26) we can assert that there exists

a positive scalar $\gamma(S)$ depending continuously on S such that

(30) $$|w|^2 \leq \gamma(S)||z||^2 \, , \quad \forall \; w \epsilon R(A), \; z\epsilon N(A)^+, \quad w = Az \, .$$

Therefore

(31) $$|y - y^*|^2 \leq \gamma(S)||x - \tilde{p}(x)||^2 \leq \gamma(S)||x - p(x)||^2 \, .$$

Combination of (29) and (31) yields (24b) for

$$r(S) = \Lambda(S) \, L^2 \, \gamma(S).$$

This completes the proof of part a).

b) Let $\bar{\alpha}$ be any positive scalar such that

$$\bar{\alpha} < \frac{2q(S)}{r(S)} \, .$$

Then using (18) we have for all $\alpha \varepsilon (0, \bar{\alpha}]$

$$(32) \qquad ||x_{k+1} - p(x_{k+1})||^2 \leq t(\alpha)||x_k - p(x_k)||^2$$

where

$$t(\alpha) = 1 - 2\alpha q(S) + \alpha^2 r(S) \quad .$$

In view of the fact $\alpha \leq \bar{\alpha} < \dfrac{2q(S)}{r(S)}$ it is easily seen that

$$0 \leq t(\alpha) < 1 \quad .$$

We have from (32)

$$(33) \qquad ||x_k - p(x_k)||^2 \leq t(\alpha)^k ||x_0 - p(x_0)||^2$$

and using (23) and (24) we obtain

$$(34) \qquad ||x_{k+1} - p(x_k)|| \leq t(\alpha)^{1/2}||x_k - p(x_k)|| \quad .$$

By the triangle inequality

$$(35) \qquad ||x_{k+1} - x_k|| \leq ||x_{k+1} - p(x_k)|| + ||x_k - p(x_k)|| \quad .$$

Combining (35) with (33), (34) we obtain

$$(36) \qquad ||x_{k+1} - x_k|| \leq [t(\alpha)^{1/2} + 1]||x_k - p(x_k)||$$

$$\leq [t(\alpha)^{1/2} + 1] \, t(\alpha)^{k/2}||x_0 - p(x_0)||.$$

Let

$$\beta = t(\alpha)^{1/2}$$

$$\bar{q} = [t(\alpha)^{1/2} + 1]||x_0 - p(x_0)|| \quad .$$

Then (36) can be written as

$$||x_{k+1} - x_k|| \leq \bar{q}\beta^k .$$

For all $k \geq 0$, $m \geq 1$ we have

(37) $$||x_{k+m} - x_k|| \leq ||x_{k+m} - x_{k+m-1}|| + .. + ||x_{k+1} - x_k||$$

$$\leq \bar{q} (\beta^{k+m-1} + ...+ \beta^k) = \frac{\bar{q}\,\beta^k(1 - \beta^m)}{1 - \beta}$$

Since in view of $t(\alpha) < 1$ we have $\beta < 1$, it follows that $\{x_k\}$ is a Cauchy sequence and hence converges to a vector x^*. Since by (33) $\{x_k - p(x_k)\}$ converges to zero we must have $x^* = p(x^*)$ which implies $x^* \epsilon X^*$. By taking limit in (37) as $m \to \infty$ we obtain for all $k = 0,1,...$

$$||x_k - x^*|| \leq q\beta^k$$

where $q = \bar{q}(1 - \beta)^{-1}$. Q.E.D.

It is quite remarkable that as shown in Proposition 1b), the sequence $\{x_k\}$ converges to a single limit point even though the set of solutions X^* may contain an infinite number of points.

3. <u>An Algorithm for the Traffic Assignment Problem</u>

In this section we consider the following network flow problem. A network consisting of a set of nodes N and a set of directed links L is given, together with a set W of node pairs referred to as <u>origin-destination (OD) pairs.</u> For each OD pair w ϵ W there is a known demand

$d_w > 0$ representing traffic entering the network at the origin and exiting at the destination. For each OD pair w, the demand d_w is to be distributed among a given collection P_w of simple directed paths joining w. We denote by x_p the flow carried by path p. Thus the set of feasible path flow vectors $x = \{x_p | p \in P_w, w \in W\}$ is given by

$$(38) \quad X = \{x | \sum_{p \in P_w} x_p = d_w, \; x_p \geq 0, \; \forall \; p \in P_w, w \in W\} \quad .$$

Each collection of path flows $x \in X$ defines a collection of link flows y_{ij}, $(i,j) \in L$ by means of the equation

$$(39) \quad y_{ij} = \sum_{w \in W} \sum_{p \in P_w} \delta_p(i,j) \, x_p \quad , \qquad \forall \; (i,j) \in L$$

where $\delta_p(i,j) = 1$ if path p contains link (i,j) and $\delta_p(i,j) = 0$ otherwise. The vector of link flows $y = \{y_{ij} | (i,j) \in L\}$ corresponding to $x \in X$ can be written as

$$y = Ax$$

where A is the arc-chain matrix defined by (39). The set of feasible link flows is thus

$$(40) \quad Y = AX = \{y | y = Ax, \; x \in X\} \quad .$$

We assume that for each link (i,j) L there is given a function $T_{ij} : Y \to R$ such that $T_{ij}(y) > 0$ for all $y \in Y$. The value of $T_{ij}(y)$ represents a measure of delay in traversing link (i,j) when the set of link flows is y (travel time in transportation networks [4], [5], marginal delay in communication networks [7], [8]). The vector with components $T_{ij}(y)$ is denoted $T(y)$. We assume that $T(y)$ is Lipschitz continuous and strongly monotone. This is a reasonable assumption for transportation networks, as well as for communication networks.

For each $x \in X$ and corresponding $y = Ax$ the vector $T(y)$ defines for each $w \in W$ and $p \in P_w$ a length

(41)
$$\bar{T}_p(x) = \sum_{(i,j)\in L} \delta_p(i,j) \, T_{ij}(y)$$

which may be viewed as the total travel time of path p. The problem is to find $x^* \in X$ such that for all $\bar{p} \in P_w$, and $w \in W$

(42)
$$\bar{T}_{\bar{p}}(x^*) = \min_{p \in P_w} \bar{T}_p(x^*) \quad , \qquad \text{if } x_{\bar{p}}^* > 0 \ .$$

This problem is based on the user-optimization principle which asserts that a traffic network equilibrium is established when no user may decrease his travel time by making a unilateral decision to change his route. If we denote by $\bar{T}(x)$ the vector of lengths $\bar{T}_p(x)$, $p \in P_w$, $w \in W$ then it is easy to see ([4], [5]) using (39) and (41) that

(43)
$$\bar{T}(x) = A'T(Ax)$$

and that the problem defined earlier via (42) is equivalent to finding a solution $x^* \in X$ of the variational inequality

(44)
$$(x - x^*)' \, \bar{T}(x^*) \geq 0, \qquad \forall \ x \in X,$$

which is of the form (1).

Let $W = \{w_1, w_2, \ldots, w_M\}$ and consider the algorithm of the previous section with a matrix S which is block diagonal of the form

$$S = \begin{bmatrix} S_1 & & & 0 \\ & S_2 & & \\ & & \ddots & \\ 0 & & & S_M \end{bmatrix} \ .$$

Here each matrix S_i corresponds to the OD pair w_i. We assume that each matrix S_i is positive definite symmetric. The projection iteration (4) can be implemented by finding x_{k+1} solving the quadratic program (16). In view of the block diagonal form of S and the decomposable nature of the constraint set X [c.f.(38)] the quadratic program can be decomposed

into a collection of smaller quadratic programs - one per OD pair $w \in W$.
The form of these programs for the case where each matrix S_i is diagonal
with elements s_p, $p \in P_{w_i}$ along the diagonal is

$$(45) \qquad \text{minimize} \sum_{p \in P_{w_i}} \{(x_p - x_p^k)\ \bar{T}_p\ (x_k) + \frac{s_p}{2\alpha}(x_p - x_p^k)^2\}$$

$$\text{subject to} \sum_{p \in P_{w_i}} x_p = d_{w_i}\ ,\quad x_p \geq 0,\quad \forall\ p \in P_{w_i}$$

where x_p^k, $p \in P_{w_i}$, $i = 1,\ldots,M$ are the components of the vector x_k. Problem (45)
involves a single equality constraint and can be solved very easily--essentially
in closed form [8], [10]. The convergence and linear rate of convergence
results of Proposition 1 apply to this algorithm.

The preceding algorithm is satisfactory if each set P_{w_i} contains
relatively few paths. In some problems however the number of paths in
P_{w_i} can be very large (for example P_{w_i} may contain all simple paths joining
w_i). In this case it is preferable to start with a subset of each P_{w_i} and
augment this subset as necessary as suggested by Aashtiani [5]. The
corresponding algorithm is as follows:

We begin with a subset $P_{w_i}^o \subset P_{w_i}$ for each $w_i \in W$ and a vector of initial
path flows x_o such that for all $w_i \in W$

$$x_p^o = 0 \qquad \text{if}\quad p \notin P_{w_i}^o\ .$$

At the kth iteration we have for each OD pair $w_i \in W$ a subset of paths $P_{w_i}^k \subset P_{w_i}$
and a corresponding vector of path flows x_k satisfying

$$x_p^k = 0 \qquad \text{if}\qquad p \notin P_{w_i}^k\ .$$

We compute, for each $w_i \in W$, a shortest path $p_{w_i}^k \in P_{w_i}$ using $T_{ij}(Ax_k)$ as length

for each link (i,j). We set

$$P_{w_i}^{k+1} = P_{w_i}^k \cup \{p_{w_i}^k\} \quad .$$

(Note here that $p_{w_i}^k$ may already belong to $P_{w_i}^k$). We then solve the quadratic programming problem

$$(46) \qquad \text{minimize} \sum_{p \in P_{w_i}^{k+1}} \{(x_p - x_p^k)\, \bar{T}_p(x_k) + \frac{s_p}{2\alpha}(x_p - x_p^k)^2\}$$

$$\text{subject to} \sum_{p \in P_{w_i}^{k+1}} x_p = d_{w_i}, \quad x_p \geq 0, \quad \forall\ p \in P_{w_i}^{k+1} \quad .$$

If $\bar{x}_p^k$, $p \in P_{w_i}^{k+1}$ is the solution of this problem we set

$$(47) \qquad x_p^{k+1} = \begin{cases} \bar{x}_p^k & \text{if } p \in P_{w_i}^{k+1} \\ \\ 0 & \text{if } p \notin P_{w_i}^{k+1} \end{cases} \quad .$$

Note that the quadratic programming problem (46) is the same as (45) except for the fact that it involves only paths in the subset $P_{w_i}^{k+1}$. The subset is possibly augmented at each iteration k to include the current shortest path $p_{w_i}^k$. The expectation here is that, while P_{w_i} may contain a very large number of paths, the actual number of paths generated and included in the set $P_{w_i}^k$ remains small as k increases. This expectation has been supported by computational experiments [5]. The convergence and rate of convergence properties of the algorithm (46), (47) are identical with those of the earlier algorithm based on (45) as the reader can easily verify. The key idea is based on the fact that for each i, P_{w_i} contains a finite number of paths and $P_{w_i}^k$ grows monotonically so that the sequence $\{P_{w_i}^k\}$ converges to some subset of paths $\bar{P}_{w_i} \subset P_{w_i}$. By applying Proposition 1 it follows that the algorithm converges to a solution of the problem obtained when P_{w_i} is replaced by $\bar{P}_{w_i}$. Because

$P_{w_1}^k$ is augmented at each iteration with the current shortest path it is

a simple matter to conclude that the solution x^* obtained satisfies the

required minimum travel time condition (42).

The results and algorithms of this section can be strengthened

considerably in the case where T is in addition the gradient of convex

function F in which case the problem is equivalent to the convex programming

problem of minimizing $F(y)$ subject to $y \in Y$ or minimizing $F(Ax)$ subject to $x \in X$.

Under these circumstances there are convergence results [8] - [10] relating

to projection type algorithms which allow for a variable matrix S and

for coordinate descent type iterations whereby each iteration is performed

with respect to a single OD pair (or a small group of OD pairs) and all OD

pairs are taken up in sequence. In fact for such algorithms it seems that

it is easier to select an appropriate value for the stepsize α. Aashtiani's

computational experience [5] suggests that such algorithms also work well

in many cases where T is not a gradient mapping. We have been unable

however to obtain a general convergence result for coordinate descent

versions of the projection algorithm. The possibility of changing the

matrix S from one iteration to the next is considered in the next section.

4. An Algorithm with Variable Projection Metric

A drawback of the algorithms of Sections 2 and 3 is that the matrix S

is restricted to be the same at each iteration. Computational experience

with optimization problems as well as network flow problems [5], [6] suggests

that, if T is differentiable, better results can be obtained if the

matrix S is varied from one iteration to the next and is made suitably

dependent on first derivatives of the mapping T in a manner which approximates

Newton's method. We have not been able to show the result of Proposition 1b)

for algorithms in which the matrix S may change arbitrarily. On the other

hand it is possible to construct an algorithmic scheme that allows for

a variable matrix S but at the same time incorporates a mechanism that

safeguards against divergence. The main idea in this scheme is <u>to allow</u>

<u>a change in the matrix S only when the algorithm makes satisfactory progress</u>

<u>towards convergence.</u> The algorithm is as follows:

A set S of positive definite symmetric matrices is given. It is

assumed that all eigenvalues of all matrices $S \in S$ lie in some compact interval

of positive real numbers, i.e. there exist $m_1, m_2 > 0$ such that

$m_1 |z|^2 \leq z' S z \leq m_2 |z|^2$,for all $z \in R^n$, $S \in S$. We consider the algorithm

$$(48) \qquad x_{k+1} = P_X^{S_k} [x_k - \alpha S_k^{-1} \bar{T}(x_k)] \qquad , \qquad x_0 \in X$$

where $S_k \in S$ for all   k = 0,1,.... The stepsize α is such that

$$(49) \qquad h \overset{\Delta}{=} \max_{S \in S} [1 - 2\alpha \, q(S) + \alpha^2 \, r(S)] < 1,$$

where $q(S)$ and $r(S)$ are as in Proposition 1a) [c.f. (18)]. The maximum in (49)

is attained by continuity of $q(\cdot)$ and $r(\cdot)$. It is clear that there

exists $\bar{\alpha} > 0$ such that (49) is satisfied for all $\alpha \in (0, \bar{\alpha}]$. Given x_{k+1} the

matrix S_{k+1} is either chosen arbitrarily from S or else

$$S_{k+1} = S_k$$

depending on whether the quantity

$$(50) \qquad w_k = (x_{k+1} - x_k)' \, S_k (x_{k+1} - x_k)$$

has decreased or not by a certain factor over the last time the matrix S was

changed. More specifically a scalar $\bar{\beta} \in (0,1)$ (typically close to unity)

is chosen, and at each iteration k a scalar $\bar{w}_{k+1}$ is computed according to

$$\bar{w}_{k+1} = \begin{cases} \beta w_k & \text{if } w_k \leq \bar{w}_k \\ \\ \bar{w}_k & \text{if } w_k > \bar{w}_k \end{cases}$$

where w_k is given by (50) and initially $\bar{w}_o = \infty$. We select

(52a) $\qquad S_{k+1} = S_k$, $\qquad$ if $\bar{w}_{k+1} = \bar{w}_k$,

(52b) $\qquad S_{k+1} \in S$, $\qquad$ if $\bar{w}_{k+1} < \bar{w}_k$.

Thus for each k, the scalar $\bar{w}_k$ represents a target value below which w_k must drop in order for a change in S to be allowed in the next iteration.

We first show that if $\{x_k\}$ is a sequence generated by the algorithm just described then

(53) $\qquad \lim_{k \to \infty} \inf w_k = 0$.

Indeed if $\lim_{k \to \infty} \inf w_k > 0$ then S_k must have been allowed to change only a finite number of times in which case it follows from Proposition 1 that $\{x_k\}$ converges to a solution x^*. As a result we have $w_k \to 0$ contradicting the earlier assertion.

Let us denote by $||\cdot||_k$ the norm corresponding to S_k and by $p_k(z)$ the projection of a vector $z \in R^n$ on X^* with respect to $||\cdot||_k$. We have by using the triangle inequality, (49), (50) and Proposition 1a)

(54) $\qquad w_k^{1/2} = ||x_{k+1} - x_k||_k$

$\qquad\qquad \geq ||x_k - p_k(x_{k+1})||_k - ||x_{k+1} - p_k(x_{k+1})||$

$\qquad\qquad \geq ||x_k - p_k(x_k)||_k - \sqrt{h} \, ||x_k - p_k(x_k)||_k$

$\qquad\qquad = (1 - \sqrt{h}) \, ||x_k - p_k(x_k)||_k$.

Hence (53) implies $\lim_{k\to\infty} \inf ||x_k - p_k(x_k)|| = 0$ or equivalently (in view of the fact $||z||_k^2 \geq m_1 |z|^2$, $m_1 > 0$)

(55) $\lim_{k\to\infty} \inf |x_k - p_k(x_k)| = 0$.

This means that at least a subsequence of $\{x_k\}$ converges to the solution set X^*. We will show that in fact for some vector $x^* \in X^*$, and some scalars $q > 0$ and $\beta \in (0,1)$ we have

$$|x_k - x^*| \leq q \beta^k \quad , \quad \forall \ k = 0,1,\ldots$$

i.e. $\{x_k\}$ converges to a solution x^* at a rate which is at least linear.

We have

(56) $w_k = ||x_k - p_k(x_k) + p_k(x_k) - x_{k+1}||_k^2$

$$\leq ||x_k - p_k(x_k)||_k^2 + ||p_k(x_k) - x_{k+1}||_k^2$$

$$+ 2||x_k - p_k(x_k)||_k \ ||p_k(x_k) - x_{k+1}||_k$$

Using (23),(24), and (49) we obtain

(57) $||p_k(x_k) - x_{k+1}||_k^2 \leq h \ ||x_k - p_k(x_k)||_k^2$,

Combination of (56) and (57) yields

(58) $w_k \leq (1 + \sqrt{h})^2 \ ||x_k - p_k(x_k)||_k^2$

Also in view of (49) and Proposition 1 we have that there exists a scalar $d \geq 1$ such that for all k

$$||x_{k+1} - p_{k+1}(x_{k+1})||_{k+1}^2 \leq ||x_{k+1} - p_k(x_{k+1})||_{k+1}^2$$

$$\leq d \ ||x_{k+1} - p_k(x_{k+1})||_k^2$$

$$\leq dh \ ||x_k - p_k(x_k)||_k^2$$

and finally using (54)

(59) $$||x_{k+1} - p_{k+1}(x_{k+1})||^2_{k+1} \leq \frac{dh}{1-\sqrt{h}} w_k.$$

Let

(60) $$K = \{k|w_k \leq \bar{w}_k\} \quad .$$

By (51), (52) we have

(61) $$S_{k+1} = S_k , \qquad \forall k \notin K .$$

Also from (51), (53) and the fact that if $w_{\bar{k}} = 0$ then $w_k = 0$ for all $k \geq \bar{k}$, it follows that K contains an infinite number of indices. Let k_1 and k_2 be two successive indices in K with $k_1 < k_2$. Then

(62) $$w_{k_2} \leq \bar{w}_{k_2} = \bar{\beta} w_{k_1}$$

while, if $k_2 - k_1 > 1$, we have

(63) $$w_{k_1+m} > \bar{w}_{k_2} = \bar{\beta} w_{k_1} , \qquad \forall m = 1,...,(k_2-k_1-1) .$$

We also have

$$S_{k_1+1} = ... = S_{k_2} \quad .$$

In the case where $k_2 - k_1 > 1$, Proposition 1a) together with (49) and (59) yields for all $m = 1, ...,(k_2-k_1-1)$

(64) $$||x_{k_1+m} - p_{k_1+m}(x_{k_1+m})||^2_{k_1+1} \leq h^{m-1}||x_{k_1+1} - p_{k_1+1}(x_{k_1})||^2_{k_1+1}$$

$$\leq \frac{d\,h^{m-1}}{1-\sqrt{h}} w_{k_1} \quad .$$

Using (58) and (64) we have

(65) $$w_{k_1+m} \leq \frac{d\,h^{m-1}(1+\sqrt{h})^2}{1-\sqrt{h}} w_{k_1} , \qquad \forall m = 1,...,(k_2-k_1-1).$$

Inequalities (63) and (65) yield

$$\bar{\beta} w_{k_1} < w_{k_2-1} \le \frac{d\, h^{k_2-k_1-2}(1+\sqrt{h})^2}{1-\sqrt{h}}\, w_{k_1} \quad .$$

It follows that if $k_2 - k_1 \ge 1$ then

$$(66) \qquad k_2 - k_1 \le 2 + \frac{\ln \frac{\bar{\beta}(1-\sqrt{h})}{d(1+\sqrt{h})^2}}{\ln h} \quad ,$$

so if $\bar{m}$ is any positive integer such that

$$(67) \qquad 2 + \frac{\ln \frac{\bar{\beta}(1-\sqrt{h})}{d(1+\sqrt{h})^2}}{\ln h} \le \bar{m}$$

we have

$$(68) \qquad k_2 - k_1 \le \bar{m}$$

for any two successive indices k_1, k_2 in K . It follows using (62) that

$$(69) \qquad w_k \le w_0\, (\bar{\beta}^{\frac{1}{\bar{m}}})^k \quad , \qquad\qquad \forall\ k \varepsilon K \quad .$$

Using (65), (68) and (69) we also obtain

$$(70) \qquad w_k \le \frac{d(1+\sqrt{h})^2\, w_0}{1-\sqrt{h}}\ (\bar{\beta}^{\frac{1}{\bar{m}}})^{k-\bar{m}} \quad , \qquad \forall\ k \notin K \quad .$$

Combining (69) and (70) we have for some scalar $\bar{q} > 0$

$$(71) \qquad w_k \le \bar{q}(\bar{\beta}^{\frac{1}{\bar{m}}})^k \quad , \qquad\qquad \forall\ k = 0,1,\ldots$$

Since

$$m_1 |x_{k+1} - x_k|^2 \le ||x_{k+1} - x_k||_k^2 = w_k$$

relation (71) yields

$$|x_{k+1} - x_k|^2 \le \frac{\bar{q}}{m_1} (\bar{\beta}^{\frac{1}{m}})^k , \qquad \forall \ k = 0,1,\ldots$$

Since $\bar{\beta} < 1$, it follows in exactly the same manner as in the proof of Proposition 1b) that $\{x_k\}$ is a Cauchy sequence which converges to a vector $x^* \in X^*$. Furthermore for some $q > 0$ and $\beta \in (0,1)$ we have

$$(72) \qquad |x_k - x^*| \le q\beta^k , \qquad k = 0,1,\ldots$$

We have thus proved the following proposition:

Proposition 2: There exists $\bar{\alpha} > 0$ such that if $\alpha \in (0,\bar{\alpha}]$, a sequence $\{x_k\}$ generated by iteration (48) with $\{S_k\}$ selected according to (50) - (52) converges to a solution x^* of the variational inequality (1). Furthermore there exist scalars $q > 0$ and $\beta \in (0,1)$ such that

$$|x_k - x^*| \le q\beta^k , \qquad \forall \ k = 0,1,\ldots$$

5. A Computational Example

In this section we report computational results for a traffic assignment problem. The corresponding network is shown in Figure 2, and may be viewed as a model of a circular highway. There are five origins and destinations numbered 1,2,3,4,5 and connected through the highway via entrance and exit ramps. We consider the five OD pairs (1,4), (2,5), (3,1), (4,2), (5,3). Each OD pair has two paths associated with it - the clockwise and counterclockwise paths on the corresponding circle. The expressions for the travel time on each link are shown in Figure 2
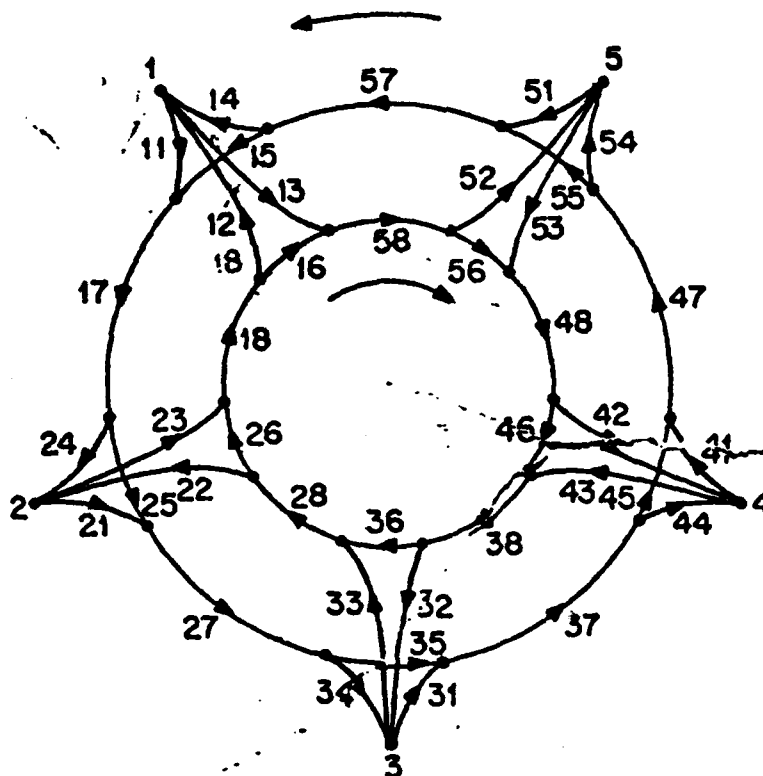
Figure 2.

Types of links:

1. **Highway Links:** 17,27,37,47,57,18,28,38,48,58

2. **Exit Ramps:** 14,24,34,44,54,12,22,32,42,52

3. **Entrance Ramps:** 11,21,31,41,51,13,23,33,43,53

4. **Bypass Links:** 15,25,35,45,55,16,26,36,46,56

Delay on Links [where g is defined by $g(x) = 1 + x + x^2$]:

1. **Delay on Highway Link** k:$10 \cdot g$[Flow on k] + $2 \cdot \gamma \cdot g$[Flow on exit ram from k]

2. **Delay on Exit Ramp** k: g[Flow on k]

3. **Delay on Entrance Ramp** k:g [Flow on k] + $\gamma \cdot g$ [Flow on Bypass Link merging with k]

4. **Delay on Bypass Link** k: g[Flow on k].

Remark: Flow is not allowed to use exit ramp not leading to its destination

OD Pairs: (1,4), (2,5), (3,1), (4,2), (5,3).

where the function g is given by

$$g(x) = 1 + x + x^2 \quad .$$

Different values of the nonnegative scalar γ represent different degrees of dependence of the travel times of some links on the flows of other links. The problem is equivalent to an optimization problem if and only if there is no such dependence ($\gamma=0$).

Tables 1 and 2 list representative computational results for two demand patterns, three values of γ, and fifteen iterations of two different algorithms labeled "all-at-once" and "one-at-a-time" and described below. The number shown in the tables for each iteration k is the following normalized measure of convergence

$$(73) \qquad \sum_{\text{all OD pairs } w} \frac{\Delta x_{w,k}}{d_w} \frac{\Delta \bar{T}_{w,k}}{\bar{T}_{min,w,k}}$$

where d_w is the demand of the OD pair w, $\Delta x_{w,k}$ is the portion of the demand that does not lie on the shortest path of the OD pair w at the end of iteration k, $\Delta \bar{T}_{w,k}$ is the difference of the travel times of the longest and shortest paths and $\bar{T}_{min,w,k}$ is the travel time of the shortest path. Clearly the expression (73) is zero if and only if the corresponding traffic assignment is optimal. The starting flow pattern in all runs was the worst possible whereby all the demand of each OD pair is routed on the counter-clockwise path. The results suggest that the algorithms yield near optimal flow patterns after very few iterations and subsequently continue their progress at a fairly satisfactory rate. This type of convergence behavior is consistent with the one observed for related algorithms tested in [6].

The "all-at-once" algorithm is the one of the previous section [c.f.(48), (50) - (52)] with the projection matrix S_k being diagonal

TABLE 1: Demands $d(1,4) = 0.1$; $d(2,5) = 0.2$; $d(3,1) = 0.3$; $d(4,2) = 0.4$; $d(5,3) = 0.5$

k	All-at-once, $\alpha = .8$, $\bar{\beta} = .99$			One-at-a-time, $\alpha = 1$, $\bar{\beta} = .99$		
	$Y = 0$	$Y = 0.5$	$Y = 4$	$Y = 0$	$Y = 0.5$	$Y = 4$
0	$.14417 \times 10^{2}$	$.14793 \times 10^{2}$	$.17426 \times 10^{2}$	$.14417 \times 10^{2}$	$.14793 \times 10^{2}$	$.17426 \times 10^{2}$
1	$.14897 \times 10^{1}$	$.15079 \times 10^{1}$	$.16436 \times 10^{1}$	$.43831 \times 10^{0}$	$.46175 \times 10^{0}$	$.36140 \times 10^{0}$
2	$.39463 \times 10^{0}$	$.36291 \times 10^{0}$	$.23633 \times 10^{0}$	$.73026 \times 10^{-1}$	$.80765 \times 10^{-1}$	$.13319 \times 10^{0}$
3	$.35901 \times 10^{0}$	$.29642 \times 10^{0}$	$.12340 \times 10^{0}$	$.28867 \times 10^{-1}$	$.37109 \times 10^{-1}$	$.39248 \times 10^{-1}$
4	$.55230 \times 10^{-1}$	$.41860 \times 10^{-1}$	$.11996 \times 10^{-1}$	$.11671 \times 10^{-1}$	$.17501 \times 10^{-1}$	$.15697 \times 10^{-1}$
5	$.80434 \times 10^{-1}$	$.52264 \times 10^{-1}$	$.55236 \times 10^{-2}$	$.50979 \times 10^{-2}$	$.86871 \times 10^{-2}$	$.10786 \times 10^{-1}$
6	$.11485 \times 10^{-1}$	$.70972 \times 10^{-2}$	$.14528 \times 10^{-2}$	$.23552 \times 10^{-2}$	$.45007 \times 10^{-2}$	$.76687 \times 10^{-2}$
7	$.19034 \times 10^{-1}$	$.93712 \times 10^{-2}$	$.71098 \times 10^{-3}$	$.11224 \times 10^{-2}$	$.23972 \times 10^{-2}$	$.55389 \times 10^{-2}$
8	$.26034 \times 10^{-2}$	$.13097 \times 10^{-2}$	$.34892 \times 10^{-3}$	$.54303 \times 10^{-3}$	$.12962 \times 10^{-2}$	$.40331 \times 10^{-2}$
9	$.43683 \times 10^{-2}$	$.21587 \times 10^{-2}$	$.18026 \times 10^{-3}$	$.26555 \times 10^{-3}$	$.70754 \times 10^{-3}$	$.29498 \times 10^{-2}$
10	$.55167 \times 10^{-3}$	$.11845 \times 10^{-2}$	$.99838 \times 10^{-4}$	$.12989 \times 10^{-3}$	$.38809 \times 10^{-3}$	$.21641 \times 10^{-2}$
11	$.10228 \times 10^{-2}$	$.11548 \times 10^{-2}$	$.80656 \times 10^{-4}$	$.64334 \times 10^{-4}$	$.21363 \times 10^{-3}$	$.15911 \times 10^{-2}$
12	$.44825 \times 10^{-3}$	$.80420 \times 10^{-3}$	$.68090 \times 10^{-4}$	$.31281 \times 10^{-4}$	$.11782 \times 10^{-3}$	$.11698 \times 10^{-2}$
13	$.51164 \times 10^{-3}$	$.67664 \times 10^{-3}$	$.59269 \times 10^{-4}$	$.15289 \times 10^{-4}$	$.64249 \times 10^{-4}$	$.86171 \times 10^{-3}$
14	$.30760 \times 10^{-3}$	$.51037 \times 10^{-3}$	$.51284 \times 10^{-4}$	$.77946 \times 10^{-5}$	$.35938 \times 10^{-4}$	$.63501 \times 10^{-3}$
15	$.27834 \times 10^{-3}$	$.41039 \times 10^{-3}$	$.44031 \times 10^{-4}$	$.41734 \times 10^{-5}$	$.19540 \times 10^{-4}$	$.46808 \times 10^{-3}$

TABLE 2: Demands $d(1,4) = 1$, $d(2,5) = 8$, $d(3,1) = 1$, $d(4,2) = 8$, $d(5,3) = 1$

k	All at once, $\alpha = .8$, $\bar{\beta} = .99$			One at a time, $\alpha = 1$, $\bar{\beta} = .99$		
	$Y = 0.0$	$Y = 0.5$	$Y = 4$	$Y = 0$	$Y = 0.5$	$Y = 4$
0	$.10203 \times 10^4$	$.10478 \times 10^4$	$.12404 \times 10^4$	$.10203 \times 10^4$	$.10478 \times 10^4$	$.12405 \times 10^4$
1	$.19446 \times 10^1$	$.15853 \times 10^1$	$.33203 \times 10^0$	$.28891 \times 10^1$	$.24262 \times 10^1$	$.86532 \times 10^0$
2	$.83731 \times 10^0$	$.57695 \times 10^0$	$.27719 \times 10^{-1}$	$.65950 \times 10^0$	$.49153 \times 10^0$	$.59564 \times 10^{-1}$
3	$.12902 \times 10^1$	$.90467 \times 10^0$	$.37155 \times 10^{-1}$	$.78641 \times 10^{-1}$	$.56482 \times 10^{-1}$	$.13636 \times 10^{-1}$
4	$.45269 \times 10^0$	$.24701 \times 10^0$	$.24096 \times 10^{-1}$	$.63059 \times 10^{-2}$	$.70762 \times 10^{-2}$	$.56463 \times 10^{-2}$
5	$.83315 \times 10^0$	$.42439 \times 10^0$	$.17847 \times 10^{-1}$	$.70066 \times 10^{-3}$	$.27832 \times 10^{-2}$	$.29389 \times 10^{-2}$
6	$.23621 \times 10^0$	$.79849 \times 10^{-1}$	$.12985 \times 10^{-1}$	$.10199 \times 10^{-3}$	$.12097 \times 10^{-2}$	$.16215 \times 10^{-2}$
7	$.50649 \times 10^0$	$.11570 \times 10^0$	$.95892 \times 10^{-2}$	$.79205 \times 10^{-4}$	$.45572 \times 10^{-3}$	$.90603 \times 10^{-3}$
8	$.11333 \times 10^0$	$.10000 \times 10^{-1}$	$.71098 \times 10^{-2}$	$.14878 \times 10^{-4}$	$.16733 \times 10^{-3}$	$.50823 \times 10^{-3}$
9	$.26874 \times 10^0$	$.78683 \times 10^{-2}$	$.52748 \times 10^{-2}$	$.94583 \times 10^{-4}$	$.61370 \times 10^{-4}$	$.28569 \times 10^{-3}$
10	$.43230 \times 10^{-1}$	$.10397 \times 10^{-2}$	$.39184 \times 10^{-2}$	$.65018 \times 10^{-4}$	$.22825 \times 10^{-4}$	$.16119 \times 10^{-3}$
11	$.11434 \times 10^0$	$.13673 \times 10^{-2}$	$.29147 \times 10^{-2}$	$.39872 \times 10^{-4}$	$.90891 \times 10^{-5}$	$.90105 \times 10^{-4}$
12	$.12722 \times 10^{-1}$	$.68513 \times 10^{-3}$	$.21694 \times 10^{-2}$	$.24651 \times 10^{-4}$	$.33388 \times 10^{-5}$	$.51215 \times 10^{-4}$
13	$.43756 \times 10^{-1}$	$.50364 \times 10^{-3}$	$.16170 \times 10^{-2}$	$.15648 \times 10^{-4}$	$.10026 \times 10^{-5}$	$.27883 \times 10^{-4}$
14	$.45550 \times 10^{-2}$	$.32649 \times 10^{-3}$	$.12047 \times 10^{-2}$	$.92416 \times 10^{-5}$	$.40281 \times 10^{-6}$	$.15870 \times 10^{-4}$
15	$.20089 \times 10^{-1}$	$.22516 \times 10^{-3}$	$.89921 \times 10^{-3}$	$.68895 \times 10^{-5}$	$.47333 \times 10^{-6}$	$.89927 \times 10^{-5}$

[c.f. (45)]. For each iteration k for which S_k was allowed to change
[c.f.(50) - (52)] the diagonal element of S_k corresponding to any one
path was taken equal to the sum of the first derivatives of the travel
times of links on that path evaluated at the kth flow x_k. This corresponds
to a diagonal approximation of Newton's method (c.f.[8], [9]). As a result
stepsizes near unity typically give satisfactory convergence behavior. In
all runs we used a stepsize α = 0.8 which is probably a bit on the high
side. The scalar $\bar{\beta}$ used in the test for allowing the matrix S_k to
change [c.f.(51)] was taken to be 0.99 in all runs. One of the most
interesting observations from our experimentation was that this test was
passed at every iteration and so the matrix S_k was changed at every iteration.

The "one-at-a-time" algorithm is similar to the "all-at-once" algorithm
described above. The only difference is that the projection is carried
out with respect to a single OD pair with flows corresponding to the other
OD pairs being kept fixed, and all OD pairs are taken up in sequence.
An iteration consists of a cycle of five projection subiterations (one
per OD pair). Each subiteration is, of course, followed by reevaluation
of the travel time of each link. Algorithms of this type resemble coordinate
descent methods for unconstrained optimization and have been suggested in
the context of network flows in [5] and [8]-[10]. The stepsize α was taken
to be unity in all runs. Also $\bar{\beta}$ was chosen to be 0.99 and again it turned
out that the matrix S_k was allowed to change in every iteration. It is
interesting to note that for this stepsize and this particular example
the algorithm tested is equivalent to an algorithm in the class proposed by
Aashtiani [5]. Tables 1 and 2 indicate a better performance for the
"one-at-a-time" algorithm. However, there is no convergence proof available

for this algorithm at present except in the case where the corresponding variational inequality is equivalent to a convex optimization problem.

Appendix : Proof of Lemma 1

We have by the Pythagorean theorem

$$||x - \tilde{p}(x)||^2 + ||\tilde{p}(x) - p(x)||^2 = ||x - p(x)||^2 \leq ||x - p[\tilde{p}(x)]||^2$$

$$= ||x - \tilde{p}(x)||^2 + ||\tilde{p}(x) - p[\tilde{p}(x)]||^2$$

Hence

(A.1) $||\tilde{p}(x) - p(x)||^2 \leq ||\tilde{p}(x) - p[\tilde{p}(x)]||^2.$

Since $p[\tilde{p}(x)]$ by definition is the unique solution of the problem of minimizing $||\tilde{p}(x) - z||^2$ over $x \in X^*$, and $p(x) \in X^*$, it follows that equality holds in (A.1) and $p(x) = p[\tilde{p}(x)]$.

Let $\tilde{p}_1(x)$ and $p_1(x)$ be the projections of x on $\tilde{X}^*$ and X^* respectively relative to the standard norm $|\cdot|$, i.e. $\tilde{p}_1(x) = \tilde{p}(x)$, $p_1(x) = p(x)$ for S equal to the identity. In order to show the existence of a continuous $\eta(S)$ such that (21) holds it will suffice to show the existence of a scalar $\eta_1 > 0$ such that for all $x \in X$

(A.2) $|x - \tilde{p}_1(x)|^2 \geq \eta_1 |p_1(x) - \tilde{p}_1(x)|^2 .$

In order to see this let $\Lambda(S)$ and $\lambda(S)$ be the largest and smallest eigenvalue of S. We have for all x, $z \in R^n$

$$\lambda(S)|x - z|^2 \leq ||x - z||^2 \leq \Lambda(S) |x - z|^2 .$$

It follows that

(A.3) $$\frac{1}{\lambda(S)}||x - \tilde{p}(x)||^2 \geq |x - \tilde{p}_1(x)|^2$$

(A.4) $$\frac{1}{\Lambda(S)}||x - p(x)||^2 \leq |x - p_1(x)|^2 \quad .$$

By the Pythagorean theorem we have

$$|p_1(x) - \tilde{p}_1(x)|^2 = |x - p_1(x)|^2 - |x - \tilde{p}_1(x)|^2$$

so (A.2) can ve rewritten as

(A.5) $$|x - \tilde{p}_1(x)|^2 \geq \frac{\eta_1}{1+\eta_1} |x - p_1(x)|^2.$$

From (A.3), (A.4), and (A.5) we obtain

(A.6) $$||x - \tilde{p}(x)||^2 \geq \gamma(S) \, ||x - p(x)||^2$$

where

$$\gamma(S) = \frac{\eta_1}{1+\eta_1} \quad \frac{\lambda(S)}{\Lambda(S)} \quad .$$

By the Pythagorean theorem we have

$$||x - p(x)||^2 = ||x - \tilde{p}(x)||^2 + ||p(x) - \tilde{p}(x)||^2$$

and by using this equation in (A.6) we obtain the desired relation

$$||x - \tilde{p}(x)||^2 \geq \eta(S) \, ||p(x) - \tilde{p}(x)||^2$$

with

$$\eta(S) = \frac{\gamma(S)}{1 - \gamma(S)} \quad .$$

We now show existence of an $\eta_1 > 0$ such that (A.2) holds for all $x\epsilon X$. By the preceding analysis this is sufficient to prove the lemma.

For each $x\epsilon X$ consider the tangent cone C_x of X at $p(x)$, i.e. the set

(A.7) $$C_x = \{z| \text{ there exists } \alpha > 0 \text{ such that } [p(x) + \alpha z]\epsilon X\}.$$

Let $\mathscr{C}$ be the collection

(A.8)
$$\mathscr{C} = \{C_x | \ x \varepsilon X\} \quad .$$

It is easily seen that $\underline{\mathscr{C} \text{ is a finite collection}}$, i.e. for some finite

set $J \subset X$ we have

(A.9)
$$\mathscr{C} = \{C_j | j \varepsilon J\} \quad .$$

Indeed since X is polyhedral it can be represented by definition ([16],

Section 19), in terms of a finite number of vectors $d_1, \ldots, d_m \varepsilon R^n$ and

scalars $b_1, \ldots, b_m$ as

$$X = \{x | d_i' x \leq b_i, \ i = 1, \ldots, m\} \quad .$$

It is easily seen that

$$C_x = \{z | \ d_i' z \leq 0, \quad \forall \ i = 1, \ldots, m \ \text{ such that } d_i' p(x) = b_i\} \quad .$$

Clearly there is only a finite number of sets of the above form. In

what follows $P_\Omega(z)$ denotes the projection of a vector $z \varepsilon R^n$ on a closed

convex set $\Omega \subset R^n$ with respect to the standard norm $|\cdot|$. The essence of

the proof of Lemma 1 is contained in the following lemma.

<u>Lemma A.1</u>: For $j \varepsilon J$ let

(A.10)
$$M_j^* = C_j \cap N(A)$$

(A.11)
$$Z_j = \{z | \ z \varepsilon C_j, \ P_{M_j^*}(z) = 0\} \quad ,$$

where N(A) is the nullspace of A. Then for each $j \varepsilon J$ there exists a scalar

$\eta_j > 0$ such that

(A.12)
$$|z - P_{N(A)}(z)|^2 \geq \eta_j |z|^2 \qquad \forall \ z \varepsilon Z_j \quad .$$

Proof: Assume the contrary, i.e. that there exist $j \varepsilon J$ and sequences $\{z_k\} \subset Z_j$, $\{\eta_j^k\} \subset R$ such that

(A.13)
$$|z_k - P_{N(A)}(z_k)|^2 < \eta_j^k |z_k|^2 , \qquad \eta_j^k \to 0.$$

We then have $z_k \neq 0$, $\forall$ k, and since both M_j^* and Z_j are clearly cones with vertex at the origin we can assume that $|z_k| = 1$, $\forall$ k. Let $\bar{z}$ be a limit point of $\{z_k\}$. We have $|\bar{z}| = 1$ and by taking limit in (A.13) we obtain

$$\bar{z} = P_{N(A)}(\bar{z})$$

i.e. $\bar{z} \varepsilon N(A)$. Since C_j is closed we also have $\bar{z} \varepsilon C_j$ and hence $\bar{z} \varepsilon M_j^*$. It follows that

(A.14)
$$P_{M_j^*}(\bar{z}) = \bar{z} \quad .$$

On the other hand since $z_k \varepsilon Z_j$ we have $P_{M_j^*}(z_k) = 0$, $\forall$ k which implies that $P_{M_j^*}(\bar{z}) = 0$. Since $|\bar{z}| = 1$, this contradicts (A.14). Q.E.D.

We now show that the desired relation (A.2) holds with

(A.15)
$$\eta_1 = \min \{\eta_j | j \varepsilon J\} > 0 \quad .$$

Choose any $x \varepsilon X$ and let $j \varepsilon J$ be such that

$$C_x = C_j \quad .$$

Let

(A.16)
$$z = x - p_1(x) \quad .$$

By a simple translation argument, the fact that $P_1(x)$ is the projection of x on $X^* \cap \{x | Ax = y^*\}$ implies that the projection of z on M_j^* is the origin so that $z \varepsilon Z_j$. A similar argument shows that

(A.17) $\qquad \tilde{p}_1(x) - p_1(x) = P_{N(A)}(z)$.

Using (A.16) and (A.17) in (A.12) we obtain

$$|x - \tilde{p}_1(x)|^2 \geq \eta_j |x - p_1(x)|^2$$

and (A.2) follows from the definition (A.15) of η_1. $\qquad$ Q.E.D.

References

[1]   Sibony, M. "Méthodes Itératives pour les Équations et Inéquations
        aux Dérivées Partielles Nonlineaires de Type Monotone", <u>Calcolo</u>,
        Vol. 7, 1970, pp.65-183.

[2]   Bakushinskij,A.B., and Poljak, B.T.,"On the Solution of Variational
        Inequalities", <u>Soviet Math. Doklady</u>, Vol. 219, 1974.

[3]   Auslender, A., <u>Optimization. Méthodes Numériques</u>, Mason, Paris, 1976.

[4]   Dafermos, S., "Traffic Equilibrium and Variational Inequalities",
        Lefschetz Center for Dynamical Systems Report, Brown University,
        to appear.

[5]   Aashtiani, H.Z., "The Multi-Model Traffic Assignment Problem",
        Ph.D. Thesis, Sloan School of Management. M.I.T., May 1979.

[6]   Bertsekas,D.P., Gafni,E.and Vastola, K.S., "Validation of
        Algorithms for Routing of Flow in Networks", <u>Proceedings of
        1978 Conference on Decision and Control</u>, San Diego, Cal., Jan.1979,
        pp. 220-227.

[7]   Gallager, R.G., "A Minimum Delay Routing Algorithm Using Distributed
        Computation", <u>IEEE Trans. on Communication</u>, Vol. COM-25, 1977, pp.73-85.

[8]   Bertsekas, D.P., "Algorithms for Nonlinear Multicommodity
        Network Flow Problems", <u>International Symposium on Systems
        Optimization and Analysis</u>, A. Bensoussan and J.L. Lions (eds.),
        Springer-Verlag, 1979, pp. 210-224.

[9]   Bertsekas, D.P., "A Class of Optimal Routing Algorithms for
        Communication Networks", <u>Proc. of 1980 ICCC</u>, Atlanta, GA., Nov. 1980.

[10]   Gafni, E.M., "Convergence of a Routing Algorithm", Laboratory for
        Information and Decision Systems Report 907, Massachusetts Institute
        of Technology, Cambridge, Mass., May 1979.

[11]   Fratta, L., Gerla, M., and Kleinrock, L., "The Flow Deviation
        Method: An Approach to Store-and Forward Communication Network
        Design", <u>Networks</u>, Vol. 3, 1973, pp. 97-133.

[12]   Cantor, D.G., and Gerla M., "Optimal Routing in a Packet Switched
        Computer Network", <u>IEEE Trans. on Computers</u>, Vol. C-23, 1974,
        pp. 1062-1069.

[13]   Dunn, J.C., "Global and Asymptotic Convergence Rate Estimates for
        a Class of Projected Gradient Processes", <u>SIAM J. on Control and
        Optimization</u>, 1981 (to appear).

[14]   Levitin, E.S., and Poljak, B.T., "Constrained Minimization Methods",
        U.S.S.R. Comp. Math. and Math. Phys., Vol. 6, 1966, pp. 1-50.

[15]   Dunn, J.C., "Newton's Method and the Goldstein Step Length Rule for
        Constrained Minimization Problems", SIAM J. Control and Optimization,
        1981 (to appear).

[16]   Rockafellar, R.T., Convex Analysis, Princeton Univ. Press,
        Princeton, N.J. 1970.

Distribution List

Defense Documentation Center Cameron Station Alexandria, Virginia 22314	12 Copies
Assistant Chief for Technology Office of Naval Research, Code 200 Arlington, Virginia 22217	1 Copy
Office of Naval Research Information Systems Program Code 437 Arlington, Virginia 22217	2 Copies
Office of Naval Research Branch Office, Boston 495 Summer Street Boston, Massachusetts 02210	1 Copy
Office of Naval Research Branch Office, Chicago 536 South Clark Street Chicago, Illinois 60605	1 Copy
Office of Naval Research Branch Office, Pasadena 1030 East Greet Street Pasadena, California 91106	1 Copy
Naval Research Laboratory Technical Information Division, Code 2627 Washington, D.C. 20375	6 Copies
Dr. A. L. Slafkosky Scientific Advisor Commandant of the Marine Corps (Code RD-1) Washington, D.C. 20380	1 Copy